基于深度残差学习的自动驾驶道路场景理解 *

宋　锐[1a]，施智平[1a†]，渠　瀛[2]，邵振洲[1b]，关　永[1b]

(1. 首都师范大学 信息工程学院 a. 成像技术北京市高精尖创新中心; b. 轻型工业机器人与安全验证北京市重点实验室，北京 100048; 2. 田纳西大学 诺克斯维尔分校工程学院，田纳西 美国 37996)

摘　要：随着道路场景理解技术的快速发展，自主驾驶领域取得了长足的进步。在相关任务中，包括道路分割、分类和车辆检测的实时性和准确性是安全性的一个关键问题。为此，提出了一个具有编码器-解码器网络结构的基于深度残差学习的方法。一方面，编码器网络结构使用不同层次的残差网络来提取高维中的抽象特征，这些特征在接下来的三个任务中共享使用；另一方面，解码器网络结构采用一种子任务的并行计算机制，即道路分割、车辆检测和道路分类任务同时执行。此外，全卷积神经网络用于对提取的图像特征进行上采样以解决道路分割问题。最终，实验结果表明在保证高精度的前提下处理帧率可达到 15 fps 以上。

关键词：道路场景理解；深度残差学习；编码器—解码器结构；全卷积网络

中图分类号：TP18　　doi: 10.3969/j.issn.1001-3695.2018.03.0234

Road scene understanding for autonomous driving via deep residual learning

Song Rui[1a], Shi Zhiping[†1a], Qu Ying[2], Shao Zhenzhou[1b], Guan Yong[1b]

(1. *a. Beijing Advanced Innovation Center for Imaging Technology, b. Beijing Key Laboratory of Light Industrial Robot & Safety Verification, College of Information Engineering, Capital Normal University, Beijing* 100048, *China*; 2. *Department of Electrical Engineering & Computer Science, The University of Tennessee, Tennessee* 37996, *USA*)

Abstract: It is making great progress in the autonomous driving field with the rapid development of road scene understanding techniques. The safety is a concerning issue with respect to the real-time and accurate performance in the related tasks which contain the road segmentation, road classification and vehicle detection. To this end, this paper proposed an approach based on deep residual learning with an encoder-decoder network structure. On the one hand, the encoder network structure useddifferent layers of residual networks to extract the **Abstract:** features in the high dimension, which shared in the next three tasks. On the other hand, the decoder network structure adopted a mechanism of parallel computing for sub-tasks, i. e. , the road segmentation, vehicle detection and road classification tasks were executed simultaneously. Additionally, it used the fully convolutional networks to upsample the extracted features to specifically solve the problem of road segmentation. At last, the experimental results show that the processing rate can effectively reach more than 15 fps with the high accuracy guaranteed.

Key words: road scene understanding; deep residual learning; encoder-decoder structure; fully convolutional networks

chinaXiv:201806.00091v1

0　引言

随着人工智能技术的快速发展，自动驾驶领域引起人们越来越多的关注，因其在日常生活中改变人们的出行方式。基于对人身安全的考虑，自动驾驶技术需要高稳定性、准确性和及时处理各种复杂的道路场景的能力。目前，深度学习技术[1]是该领域的主流方法，它已经广泛应用于道路分割、分类和车辆检测任务中，以增强自动驾驶车辆对于驾驶场景的理解能力（图1）。因此，如何更加快速准确地进行道路场景理解在自动驾驶领域具有十分重要的研究意义。目前针对以上提及的三类道路场景理解任务，典型的解决方法如下：

a) 道路分割任务。道路分割任务作为语义分割任务的一种

收稿日期：2018-03-30；**修回日期：**2018-05-14　**基金项目：**国家自然科学基金资助项目（61702348，61772351，61572331，61472468，61602325）；国家科技支撑计划资助项目（2015BAF13B01）；国际科技合作计划项目（2011DFG13000）；北京市科委项目（Z141100002014001）；北京市属高等学校创新团队建设与教师职业发展计划项目（IDHT20150507）

作者简介：宋锐（1993-），女，河北承德人，硕士研究生，主要研究方向为计算机视觉；施智平（1974-），男（通信作者），教授，主要研究方向为定理证明（shizp@cnu.edu.cn）；渠瀛（1985-），女，博士，主要研究方向为图像处理；邵振洲（1985-），男，副研究员，主要研究方向为图像处理、医疗机器人；关永（1966-），男，教授，博导，主要研究方向为形式化验证.

ChinaXiv合作期刊

chinaXiv:201806.00091v1

应用到自动驾驶场景当中的任务。Long 等人[2]提出使用深度神经网络结构解决道路分割问题，以全卷积神经网络首次实现了端到端的语义分割任务，随后结合全卷积神经网络结构。Paszke 等人[3]提出了一种编码器—解码器网络模型，利用神经网络进行图像特征提取，以提高算法泛化能力，提高了网络的运行速度和分割任务的时效性。



图 1　自动驾驶中的道路场景理解示意图

在众多典型分割方法中，常采用 VGG 网络[4]进行特征提取任务，其中 SegNet[5]、MultiNet[6]网络即采用该网络结构进行图像高维抽象特征提取任务，以完成道路分割任务，达到了良好的运行速度和准确率。

b) 车辆检测任务。Ren 等人[7]提出的基于区域推荐的方式，改进传统方法中采用滑动窗口所带来的大规模计算问题，首先使用区域推荐网络进行多个检测物体候选框的生成，接着通过不同的神经网络模型进行训练提高置信度，最终以最大置信度为检测的最终结果。另外，Redmon 等人[8]提出的改革区域推荐式目标检测框架，将全图划分为 SXS 的格子，采用一次性预测所有格子中所含目标的候选框，做到了端到端的实时目标检测。

c) 道路分类任务。自 Krizhevsky 等人[9]提出了 AlexNet 网络结构，将神经网络应用到分类任务取得了突破性进展后，深度神经网络迅速发展。在 ILSVRC 挑战赛中，涌现了许多网络结构复杂层次丰富的网络结构，如 VGG、GoogleNet[10]等网络结构。而在 2015 年，由 He 等人[11]在原始网络结构的基础上提出的深度残差网络，首次提出残差概念，采用块结构管理网络层数，针对网络层数过高而产生的网络过拟合问题有了极大的改善，并且考虑了由于卷积下采样中丢弃的图像低维特征对于分类任务的影响，极大程度地提高了物体分类的准确率。

此外，针对神经网络当中至关重要的特征提取环节，常采用结构简单整齐的 VGG 网络结构进行图像高维特征的提取任务。然而 VGG 网络结构存在一些不足，因其网络的大规模参数导致运行速度降低，不能达到自动驾驶中的实时的应用效果。

针对上述问题，本文中采用了一种典型的编码器—解码器的网络进行自动驾驶中的道路场景理解任务的解决。首先，编码器结构采用深度残差网络（ResNet）提取图像特征，深度残差网络引入 Shortcut 连接结构，使图像低维度的特征更好的和高维度的特征融合，在提高深度的同时大大提高了准确率，而解码器结构利用提取到的特征结合不同的子任务同时完成道路分割、道路分类和车辆检测任务。最后，在 KITTI 数据集上[12]进行实验和训练，通过对比不同网络层数以及不同的网络结构最终将运行速度提高至 15 fps 以上，极大地提高了图像处理的运行速度，改善了汽车对道路环境的感知能力，进而保证了自动驾驶技术的稳定性、准确性和时效性。

1　编码器—解码器网络结构

编码器—解码器网络结构可以充分地利用图像的深层次以及浅层的显著特征，通过结合深浅层次的特征，以提高任务的准确率。本文中编码器部分即为通过将图像输入含有复杂卷积结构的神经网络进行图像的特征提取，以提取图像深层次的抽象特征[13]，该部分提取的图像特征可以共享给多个子任务；而解码器部分则是通过连接相应的特定任务进行任务处理。本文的网络结构以及编码器和解码器的重要层次的具体层次输出以及参数设置如图 2 所示，较好地完成了道路分割任务以及车辆检测和道路分类任务。

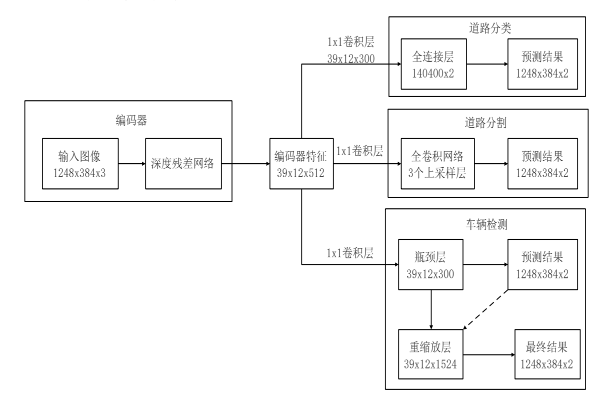


图 2　道路场景理解网络结构示意图

2　基于深度残差学习的特征提取

基于深度卷积神经网络可以从大规模的训练数据中获得复杂的更深维度的图像特征的强大优势。本文采用在 ILSVRC& COCO 2015 挑战赛中取得冠军的深度残差网络结构。区别于以往的神经网络结构，该网络引入残差学习模块 Shortcut 连接模块，在原始卷积的基础上，通过在层与层之间的输入和输出之前引入一个线性连接，这样不仅可以有效地避免因层数过多而引发的过拟合问题，同时可以更好地利用低维度的图像特征，有效地提高了准确率，如图 3 所示。

除此以外，采用 3x3 的标准卷积核，使用 ReLu 激活函数进行激活，其中包含典型的卷积以及最大池化操作。其模型包含 50、101 层，最大多达 152 层。相比 VGG 网络结构而言，减少了网络模型参数，并且加入了残差分支，使用块结构进行管理网络层数。除此以外，该网络具有强大的迁移能力可以很好地完成包括道路分类、车辆检测以及道路分割等多种任务，并且可以根据不同任务以及训练数据量的大小采用不同的网络层数的网络结构。因此，本文采用预训练的残差网络进行图像特征提取任务。表 1 中给出了本实验中所采用的不同网络结构的详细的卷积层的详细参数配置情况。因参数众多，本实验均

ChinaXiv合作期刊

chinaXiv:201806.00091v1

采用在预训练模型的基础上进行调优的做法进行网络参数的微调，以优化模型对特定数据集的适应能力。

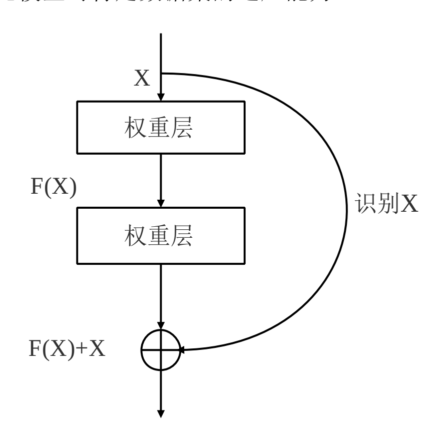


图 3 Shortcut 连接网络结构示意图

表 1 神经网络参数设置

	VGG	ResNet_50	ResNet_101	ResNet_152
Conv 1_x	[3x3,64] [3x3,64]	[7x7,64]	[7x7,64]	[7x7,64]
Conv 2_x	[3x3,128] [3x3,128]	[1×1,64 3×3,64 1×1,256] ×3	[1×1,64 3×3,64 1×1,256] ×3	[1×1,64 3×3,64 1×1,256] ×3
Conv 3_x	[3x3,256] [3x3,256] [3x3,256]	[1×1,128 3×3,128 1×1,512] ×4	[1×1,128 3×3,128 1×1,512] ×4	[1×1,128 3×3,128 1×1,512] ×8
Conv 4_x	[3x3,512] [3x3,512] [3x3,512]	[1×1,256 3×3,256 1×1,1024] ×6	[1×1,256 3×3,256 1×1,1024] ×23	[1×1,256 3×3,256 1×1,1024] ×36
Conv 5_x	[3x3,512] [3x3,512] [3x3,512]	[1×1,512 3×3,512 1×1,2048] ×3	[1×1,512 3×3,512 1×1,2048] ×3	[1×1,512 3×3,512 1×1,2048] ×3

3 基于自动驾驶的道路场景理解

3.1 基于全卷积神经网络的道路分割

全卷积网络结构作为语义分割领域的关键性进展工作，首次实现了图像端到端的语义分割任务。其提出了与卷积操作逆向的运算思路，在特征提取进行卷积下采样丢弃了图像的低维度的多种特征的劣势的前提下，将经过残差网络训练的特征经过 1x1 的卷积层重新调整维度以适应分割任务。引入上采样的操作，避免了由于使用像素块而带来的重复存储和计算卷积的问题，其采用的方法正与卷积操作相反，采用反卷积的方式来完成上采样操作。引入连接跳跃层，将低维度特征与高维度特征有机的进行结合。另外通过与条件随机场结合引入膨胀卷积结构[14]，可以在不减少维度的前提下增大感受野的范围，得到更加准确的分割结果。

3.2 基于推荐的车辆检测

车辆检测任务主要借鉴了基于推荐的方法，主要采用得益于 YOLO[8]等成功模型的 FastBox 方式，使用感兴趣区域池化的方法，充分利用深度残差网络训练所得的高维度特征。类似于分割任务，首先需要将编码器特征通过一个 1x1 的卷积层来调整网络维度，紧接着通过一个瓶颈层，该瓶颈层由多个 1x1 的卷积组成，将输出调整为 6 通道。其中前两通道表示该检测物体语义含义，数值表示其在边界框中的置信度；后四个通道表示其边界框的坐标和尺寸。这样就得到了一个粗略的估计结果。然而这种预测是不准确的，因此本文又引入了重缩放层，该层通过利用除了最大值抑制被选出的边界框以外的图像其他区域的高维特征和隐含特征，修正原始预测结果。经过感兴趣区域池化的方式，最终通过 1x1 卷积调整维度，得到最终的检测结果。

3.3 基于全连接结构的道路分类

本文针对道路分类问题，采用典型的神经网络结构中的全连接层结构。首先将经过残差网络训练的特征经过 1x1 的卷积调整图像维度，利用多分类器，使用 softmax 激活函数的全连接层结构，使用 one-hot 编码方式根据最终比例分数得到最终的预测分类的结果。

4 实验结果与分析

为了评估基于深度残差网络的自动驾驶道路场景理解算法的性能，本文进行了两组实验。第一组实验主要验证本算法自身的通用性和解决实际问题的必要性；在第二组实验中，将本文算法与同样解决道路分割的 ENet[3]、FCN[2]、SPL[15]以及进行车辆检测的 KITTI 排行榜中的算法进行了性能的对比。



图 4 KITTI Road 数据集原始图

本文实验主要采用自动驾驶领域数据内容丰富的 KITTI 数据集[12]。其中，针对道路分割和分类方法使用 KITTI Road 数据集[16]进行评估，该数据集包括 289 张训练数据和 290 张测试数据。图 4 中展示了 KITTI Road 数据集中的原始数据图像，主要包含单车道线、多车道线以及无车道线三种类型的道路图像。其中第一行为单车道线数据，第二行为多车道线数据，第三行为无车道线数据，该数据集共包含以上三种类型的数据。针对车辆检测任务采用 KITTI Object 数据集进行训练评估，检测物体被分为容易、中等和困难三个等级。分割任务使用最大 F1 值[16]和平均准确率作为评价指标进行评估，而检测任务以检测这三类物体的平均准确率为评价标准，分类任务采用平均准确率进行评估。本实验的机器配置见表 2。

ChinaXiv合作期刊

表 2　实验环境配置

操作系统	Ubuntu
CPU	32 Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz
GPU	NVIDIA GeForce TitanX
RAM	64GB
语言环境	Python

4.1　神经网络参数设置

4.1.1 损失函数

主要包含道路分割、分类以及车辆检测任务的损失函数。因道路分割和分类任务采用同种损失函数，这里以分割任务为例，分割任务采用交叉熵作为损失函数，其定义如式（1）所示。

$$\mathrm{Loss}_{seg}(p,q)=-\frac{1}{I}\sum_{i\in I}\sum_{c\in C}q_i(c)\log p_i(c) \tag{1}$$

其中：p 是预测值；q 为实际标定值；c 为所属类别集合；I 为最小批次中成员。

针对车辆检测的损失函数采用置信度的交叉熵和边界框的坐标 L1 损失加和组成，定义如下：

$$\mathrm{Loss}_{bbox}(p,q)=-\frac{1}{I}\sum_{i\in I}\delta_{q_i}\left(\left|x_{p_i}-x_{q_i}\right|+\left|y_{p_i}-y_{q_i}\right|+\left|w_{p_i}-w_{q_i}\right|+\left|h_{p_i}-h_{q_i}\right|\right) \tag{2}$$

其中：p 是预测值；q 为实际标定值；I 为最小批当中的成员。边界框主要包含四个参数、边界框的中心点坐标(x,y)以及宽度w和高度h。

4.1.2 初始化

编码器阶段采用预先在 ImageNet 上训练过的深度残差网络权重进行初始化。车辆检测解码器权重采用随机初始化方式初始化权重，分割解码器权重采用残差网络权重进行初始化，其中包含的跳跃连接采用随机初始化的方式进行初始化。

4.1.3 优化器和正则化

本文中神经网络训练采用 Adam[17]优化器，以学习率为 1e-5 进行训练，随机失活百分比采用 0.5，所有层次权重衰减采用 5e-4。

4.2　本文提出方法的性能评估

为了探究使用深度残差网络对于自动驾驶道路场景理解任务中对于性能的影响，本文在同一数据集上对采用不同网络结构对道路分割、车辆检测以及道路分类任务进行了实验，针对道路分割任务采用最大 F1 值（MaxF1）作为对比指标，对于车辆检测任务以其中等难度的物体检测平均准确率进行比较，而分类平均准确率（AP）作为分类问题的评估标准，见表 3。

从表 3 中可以看到分别采用 VGG 网络结构和深度残差网络进行特征提取任务的结果。针对道路分割任务，相较于采用 VGG 网络结构而言，使用深度残差网络进行特征提取的分割的准确率提高到了 6.5%，对于车辆检测任务，其平均准确率也提高了 2.15%。另外，对于传统的分类任务而言，平均准确率也有了小幅的提高。实验结果进一步证明了对于自动驾驶中的道路场景理解任务而言，深度残差网络相较于 VGG 网络有利于任务准确率的提高。

表 3　不同网络结构进行道路场景理解任务的对比

	VGG	ResNet
道路分割	95.13%	96.05%
车辆检测	84.39%	86.54%
道路分类	94.38%	95.43%

表 4　不同网络层数进行道路分割任务的对比

	帧率(fps)	耗时(msec)	MaxF1(%)	AP(%)
VGG	6.59	151.74	95.13	92.32
ResNet_50	15.11	66.19	96.05	92.15
ResNet_101	9.72	102.86	95.88	92.17
ResNet_152	7.1	140.85	95.59	92.25

对于道路分割任务，从表 4 中可以看出，使用深度残差网络并采用不同层数的网络结构在分割任务中有明显的提升。实验结果将运行帧率提高到了 15.11 fps，这与使用 VGG 网络结构进行特征提取任务而言提高了 8.52 fps。另外观察使用不同层数的深度残差网络，并未达到层数越深结果越优的预测结果，笔者猜测这是由于 KITTI 数据集当中道路的数据集数量有限，对于越大层数的网络结构而言，会随着训练数据量的减小，提高其过拟合的可能性。针对准确率的评估指标最大 F1 值以及平均准确率而言，深度残差网络也在准确性上有小幅的提升，但提升效果比明显，这与网络层数的不断加深所伴随的网络模型参数大幅度提高有一定的关联。图 5 对采用不同网络结构以及不同网络层数的道路分割任务的分割评估结果进行了可视化。可以直观地发现采用深度残差网络进行道路分割任务的显著提升部分。

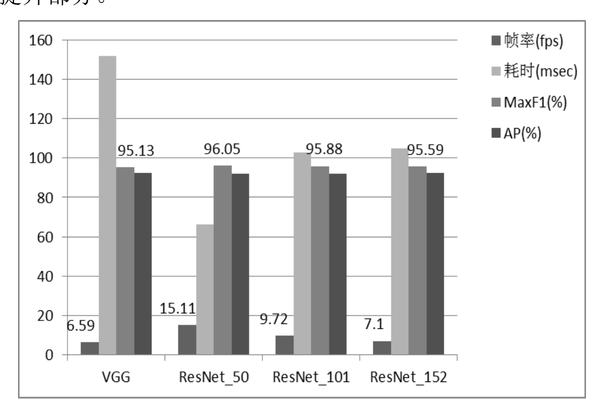


图 5　不同网络层次的道路分割柱状图

表 5　不同网络层数进行车辆检测任务的对比

	容易/%	中等/%	困难/%	帧率/fps	耗时/ms
VGG	94.2	84.5	69.7	16.530	60.496
ResNet_50	94.8	86.5	72.4	15.275	65.465
ResNet_101	96.9	89.3	75.1	9.76	102.49
ResNet_152	97.1	89.4	77.3	8.01	125.05

针对车辆检测任务，本文采用不同网络结构以及不同网络层数进行图像特征提取任务用以完成车辆检测任务。如表 5 以及图 6 所示，物体检测等级分为容易、中等以及困难三个类别。对以上三种不同类型的物体分别采用 VGG 网络和不同层数的深度残差网络结构进行训练和评估，对比结果显示，在保证运行速度的同时，本文采用的深度残差网络结构在车辆识别准确

chinaXiv:201806.00091v1

ChinaXiv合作期刊

率方面有了明显的提高，此三类以中等作为最终评估指标，152 层的深度残差网络将识别准确率提高了 4.9%。笔者猜测因 KITTI Object 数据集当中含有多种类，丰富的数据集作为预训练的数据，使得多层次、大规模的神经网络网络模型得到了充分的训练，因而识别的准确率得到了大幅的提升。

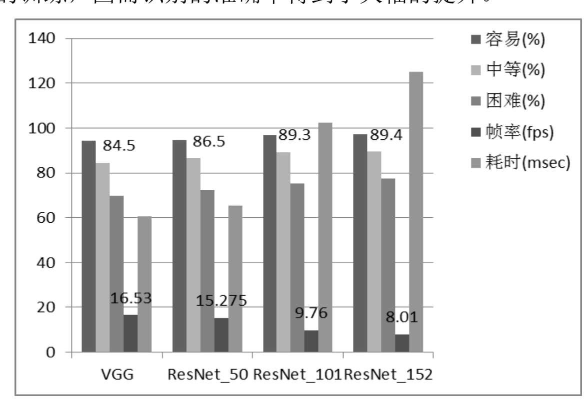


图 6　不同网络层次的车辆检测柱状图

4.3　不同道路分割以及车辆检测方法的对比

实验中分别将同样用于进行处理道路分割任务的 ENet 、FCN、 SPL 以及本文算法进行比较，其中 ENet 网络采用编码加/解码的网络，将分类反向传播给原始图像进行语义分割。FCN 网络是首个实现端到端语义分割的典型网络结构。此外，SPL 则引入了无监督的方式进行标签生成最终完成道路分割任务。表 6 对于不同方法在道路分割任务上的准确率进行了比较。

表 6　不同道路分割方法的准确率对比

	MaxF1/%	AP/%
ENet	93.13	93.01
FCN	90.89	82.32
SPL	93.69	92.96
本文方法	96.05	92.15

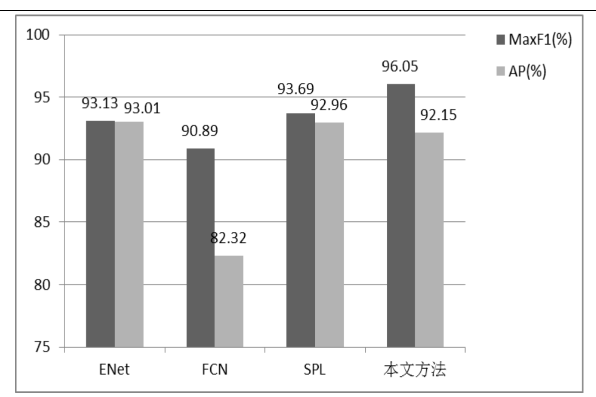


图 7　不同分割算法的准确率柱状图

从表 6 和图 7 中可以看出，使用了深度残差编码—解码的网络结构进行道路分割的方法比其他未使用没有使用深度残差编码—解码的网络结构的方法（ENet、FCN、SPL）的分割准确率有明显的提升，本文方法达到了最好的分割准确率。相较于传统的语义分割方法 FCN，本文方法的准确率提高了 5.16%。因为在处理分割任务的同时，采用编/解码的结构，并采用深度残差网络进行特征提取，将图像的高维抽象特征与低维的边界纹理特征进行深度融合，提高了网络模型的泛化能力，进而提高了分割的准确率。另外，本文仅在 KITTI 道路数据集进行训练评估，并未借助其他更大规模的数据集，这与方法 SPL 借助 KITTI Object 数据集进行训练模型相比较，节省了大量的训练时间以及数据集资源。

表 7　不同车辆检测方法的速度对比

	容易/%	中等/%	困难/%	耗时/ms	环境
UI	89.6	87.3	71.2	400	GPU@2.5 Ghz
TWSNet	90.0	86.3	71.4	480	GPU@3.5 Ghz
VCTNet	89.4	86.0	75.9	180	GPU@3.5 Ghz
本文方法	94.8	86.5	72.3	65	GPU@2.5 Ghz

针对车辆检测任务，将本文方法与 KITTI Object 排行榜中提出的不同的优秀车辆检测方法进行比较。进行检测任务比较的算法在硬件运行环境等条件下与本文方法基本一致，因此将检测结果进行比较。表 7 对于不同的检测算法，在准确率相近的前提下进行运行速度的比较。

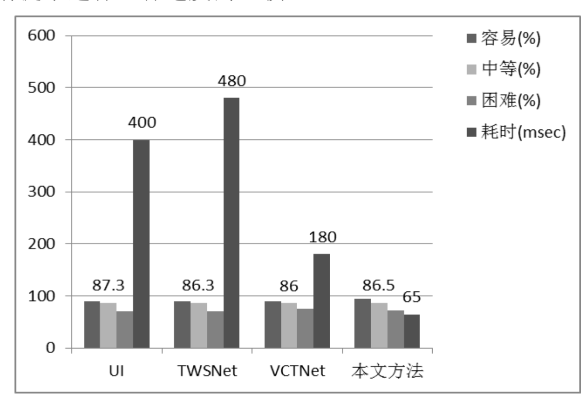


图 8　不同检测算法的性能比较柱状图

从表 7 和对应的图 8 中可以看出，将本文方法与硬件运行环境一致甚至硬件环境更优越的检测算法进行比较，在保证检测准确率无较大差距的前提下，本文算法在速度上有明显的提升；在保证较高准确率的前提下，运行时间达到了 65 ms。由于本文是在深度残差网络载入预训练模型的前提下，仅在 KITTI 数据上进行参数调优操作，所以这在缩短运行时间、提高运行速度方面有很大的提升。

图 9 以直观的方式，将道路分割任务结果进行可视化。其中，第一行中的阴影区域标记出了算法输出的道路分割区域的结果；第二行为原图中的道路的实际有效面积，道路区域为道路的实际区域；第三行为 KITTI Road 数据集当中的实际标签所显示的道路的标注区域。图 10 展示了道路分类以及车辆检测的结果。其中第一行为 KITTI Road 数据集中单车道线以及多车道线道路的原始图像；第二行中图像左上角展示图像所属道路的类别，车辆采用边界框标出本算法所检测到的车辆位置，被框出的区域为检测到的车辆的位置。结果表明，本文方法可以有效地完成道路分割、车辆检测以及道路分类任务。

针对道路分割这一特定任务，目前由奔驰主推的 Cityscapes 数据集[18]同样提供了自动驾驶环境下的图像分割数据集，用于评估视觉算法在城区场景语义理解方面的能力。同时，它提供了 50 个城市不同场景、不同季节的 5 000 张精细标注的图像，

chinaXiv:201806.00091v1

ChinaXiv合作期刊
chinaXiv:201806.00091v1

是目前自动驾驶环境下标注十分完备的数据集。

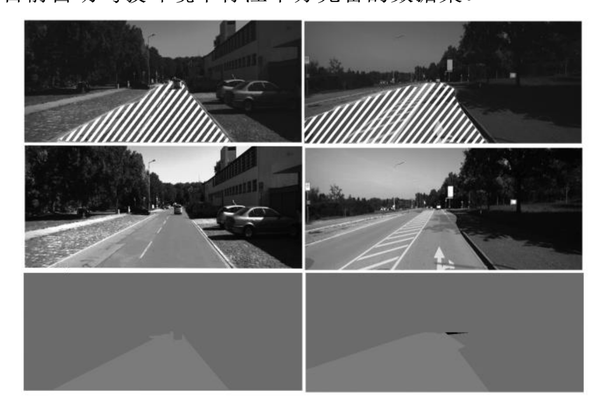


图 9　KITTI Road 道路分割结果展示图



图 10　道路分类和车辆检测结果

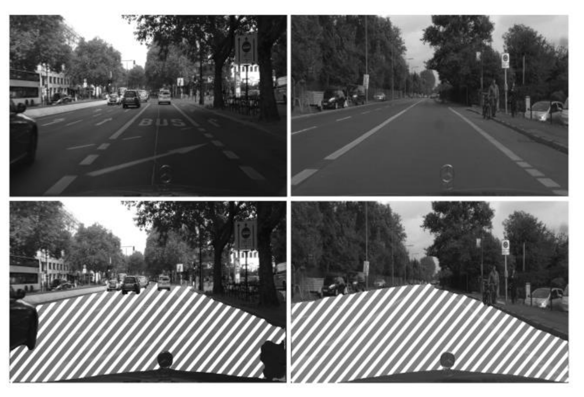


图 11　Cityscapes 数据集道路分割测试结果

基于以上实验的基础下，本文同时采用该数据集在深度残差网络下进行训练测试，提取道路特征并完成道路分割任务。在该数据集下的测试结果如图 11 所示。由图 11 可以直观地观察到深度残差网络在不同道路场景数据集下的道路分割任务的实际效果。区别于 KITTI 数据集，Cityscapes 数据集标注了 30 种不同的物体，在本实验中针对特定的道路分割任任务，仅进行道路特征的学习，将其他多余的特征作为背景进行处理。其中第一行为 Cityscapes 数据集当中的原始数据图像；第二行为使用本文当中的残差网络进行道路分割的分割结果图像，可以明显观察到在使用不同的数据集进行测试当中，本文方法也具有很好的泛化能力与使用价值，同理可以将本文方法很好地迁移到自动驾驶场景下的其他标注完备的数据集下进行测试。

5　结束语

本文针对自动驾驶领域中的道路场景理解问题，提出了基于深度残差学习的编码器—解码器网络结构用于解决相关道路场景理解问题的方法。该方法将深度残差网络作为编码器进行图像高维特征提取任务，并将提取的高维特征共享给并行的道路分割、车辆检测以及道路分类问题中，以提高运行速度和任务准确率。在 KITTI 数据集上的实验表明，该算法能够在保证道路分割精度的情况下有效提高道路分割的运行速度，并且在一定程度上提高了车辆检测以及道路分类任务的准确率。该算法改善了汽车对道路环境的感知能力，进而保证了自动驾驶技术的稳定性、准确性和时效性，在自动驾驶领域具有广泛的应用场景。

参考文献：

[1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proc of International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2012: 1097-1105.

[2] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[3] Paszke A, Chaurasia A, Kim S, *et al.* Enet: a deep neural network architecture for real-time semantic segmentation [J]. arXiv: 1606. 02147, 2016.

[4] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv: 1409. 1556, 2014.

[5] Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2017, 39 (12): 2481-2495.

[6] Teichmann M, Weber M, Zoellner M, *et al.* Multinet: real-time joint semantic reasoning for autonomous driving [EB/OL]. (2016) . https://arxiv.org/pdf/1612. 07695.pdf.

[7] Ren Shaoqing, He Kaiming, Girshick R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks [C]// Advances in Neural Information Processing Systems. 2015: 91-99.

[8] Redmon J, Divvala S, Girshick R, *et al.* You only look once: unified, real-time object detection [C]// Proc of IEEE Conference on Computer Cision and Pattern Recognition. 2016: 779-788.

[9] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proc of International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2012: 1097-1105.

[10] Szegedy C, Liu Wei, Jia Yangqing, *et al.* Going deeper with convolutions [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.] : IEEE Computer Society, 2015: 1-9.

[11] He Kaiming, Zhang Xiangyu, Ren Shaoqing, *et al.* Deep residual learning for image recognition [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

ChinaXiv合作期刊

[12] Geiger A. Are we ready for autonomous driving?The KITTI vision benchmark suite [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.] : IEEE Computer Society, 2012: 3354-3361.

[13] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks [C]// Proc of European Conference on Computer Vision. [S. l.] : Springer, 2014: 818-833.

[14] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [J]. arXiv: 1511. 07122, 2015.

[15] Wang Weiyue, Wang Naiyan, Wu Xiaomin, *et al*. Self-paced cross-modality transfer learning for efficient road segmentation [C]// Proc of IEEE International Conference on Robotics and Automation. 2017: 1394-1401.

[16] Fritsch J, Kuhnl T, Geiger A. A new performance measure and evaluation benchmark for road detection algorithms [C]// Proc of International IEEE Conference on Intelligent Transportation Systems. [S. l.] : IEEE Press, 2014: 1693-1700.

[17] Kingma D, Ba J. Adam: a method for stochastic optimization [EB/OL]. (2014) . https://arxiv. org/pdf/1412.6980. pdf.

[18] Cordts M, Omran M, Ramos S, *et al.* The cityscapes dataset for semantic urban scene understanding [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.] : IEEE Computer Society, 2016: 3213-3223.

chinaXiv:201806.00091v1